(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平10－116094

(43)公開日　平成10年(1998)５月６日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ１０Ｌ　3/00	５６１	Ｇ１０Ｌ　3/00	５６１Ｇ
	５３１		５３１Ｃ

審査請求　未請求　請求項の数32　ＯＬ　（全 12 頁）

(21)出願番号　特願平9－265959

(22)出願日　平成９年(1997)９月30日

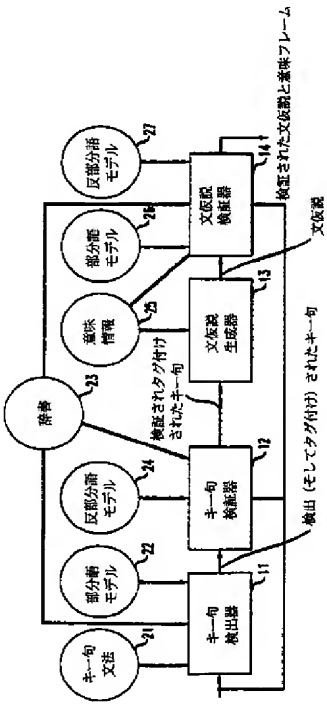
(31)優先権主張番号　０８／７２４４１３
(32)優先日　1996年10月１日
(33)優先権主張国　米国（ＵＳ）
(31)優先権主張番号　０８／７７１７３２
(32)優先日　1996年12月20日
(33)優先権主張国　米国（ＵＳ）

(71)出願人　596077259
ルーセント　テクノロジーズ　インコーポレイテッド
Ｌｕｃｅｎｔ　Ｔｅｃｈｎｏｌｏｇｉｅｓ　Ｉｎｃ．
アメリカ合衆国　07974　ニュージャージー、マレーヒル、マウンテン　アベニュー　600－700
(72)発明者　ウー　チョウ
アメリカ合衆国、07922　ニュージャージー、バークレー　ハイツ、グリーンブライアー　ドライブ　22
(74)代理人　弁理士　三俣　弘文

最終頁に続く

(54)【発明の名称】　音声認識方法および音声認識装置

(57)【要約】
【課題】　固定した形式的な文法に基づいて受容されるよりも多くの発話が受容される、効率および自由度の高い会話音声認識を実現する。
【解決手段】　キー句検出器１１で、会話の状態に固有の句部分文法の集合に基づいて複数のキー句を検出する。次に、キー句検証器１２で、これらのキー句に信頼性尺度を割り当て、その信頼性尺度をしきい値と比較することによってそれらのキー句を検証し、検証済みキー句候補の集合を得る。次に、文仮説生成器１３で、検証済みキー句候補を、所定の（例えばタスク固有の）意味情報２５に基づいて結合して文仮説を得る。最後に、文仮説検証器１４で、これらの文仮説を検証して、検証済み文仮説を生成し、最終認識結果を得る。会話ベースのシステムでは、大規模なタスク内の会話の特定の状態に応じて（すなわち「サブタスク」に対して）、個別の句部分文法を使用することが可能である。



【特許請求の範囲】
【請求項１】　複数の単語からなる音声発話の音声認識を実行する音声認識方法において、
句部分文法に基づいてキー句検出を実行して、認識された単語からなる複数の検出済みキー句を生成する検出ステップと、
前記検出済みキー句に信頼性尺度を割り当て、該信頼性尺度をしきい値と比較することにより、前記検出済みキー句の検証を実行して、検証済みキー句候補の集合を生成するキー句検証ステップと、
前記検証済みキー句候補を結合し、所定の意味情報に基づいて文仮説を生成するステップと、
前記文仮説の検証を実行して、少なくとも１つの検証済み文仮説を生成する文仮説検証ステップとからなることを特徴とする音声認識方法。
【請求項２】　前記句部分文法は、会話状態に基づく句部分文法の集合から選択されることを特徴とする請求項１の方法。
【請求項３】　前記句部分文法は、音声サンプルのコーパスを用いたトレーニングプロセスに基づいて導出されたものであることを特徴とする請求項１の方法。
【請求項４】　前記文仮説の生成は、前記信頼性尺度にも基づくことを特徴とする請求項１の方法。
【請求項５】　前記検出済みキー句は、意味タグでラベルされることを特徴とする請求項１の方法。
【請求項６】　前記文仮説の生成は、前記意味タグにも基づくことを特徴とする請求項５の方法。
【請求項７】　前記文仮説の生成は、前記信頼性尺度にも基づくことを特徴とする請求項６の方法。
【請求項８】　前記文仮説の生成は、前記信頼性尺度、前記意味タグ、および前記所定の意味情報に基づいて、最も確からしい文仮説を判定するステップからなることを特徴とする請求項７の方法。
【請求項９】　前記検出ステップは複数の部分語モデルに基づいて実行され、前記検出済みキー句は部分語の列からなることを特徴とする請求項１の方法。
【請求項１０】　前記部分語モデルは隠れマルコフモデルからなることを特徴とする請求項９の方法。
【請求項１１】　前記キー句検証ステップは、部分語モデルの集合と、対応する反部分語モデルの集合に基づいて実行されることを特徴とする請求項９の方法。
【請求項１２】　前記部分語モデルおよび前記反部分語モデルは隠れマルコフモデルからなることを特徴とする請求項１１の方法。
【請求項１３】　前記文仮説検証ステップは、文仮説に対して音響的検証を実行するステップからなることを特徴とする請求項１の方法。
【請求項１４】　前記文仮説検証ステップは、文仮説に対して意味的検証を実行するステップからなることを特徴とする請求項１の方法。
【請求項１５】　前記文仮説検証ステップは、最も確からしい１つの文仮説を選択するステップを含むことを特徴とする請求項１の方法。
【請求項１６】　前記検証済み文仮説に基づいて意味フレームを生成するステップをさらに有することを特徴とする請求項１の方法。
【請求項１７】　複数の単語からなる音声発話の音声認識を実行する音声認識装置において、
句部分文法に基づいてキー句検出を実行して、認識された単語からなる複数の検出済みキー句を生成するキー句検出器と、
前記検出済みキー句に信頼性尺度を割り当て、該信頼性尺度をしきい値と比較することにより、前記検出済みキー句の検証を実行して、検証済みキー句候補の集合を生成するキー句検証器と、
前記検証済みキー句候補を結合し、所定の意味情報に基づいて文仮説を生成する文仮説生成器と、
前記文仮説の検証を実行して、少なくとも１つの検証済み文仮説を生成する文仮説検証器とからなることを特徴とする音声認識装置。
【請求項１８】　前記句部分文法は、会話状態に基づく句部分文法の集合から選択されることを特徴とする請求項１７の装置。
【請求項１９】　前記句部分文法は、音声サンプルのコーパスを用いたトレーニングプロセスに基づいて導出されたものであることを特徴とする請求項１７の装置。
【請求項２０】　前記文仮説生成器は、前記信頼性尺度にも基づいて前記文仮説を生成することを特徴とする請求項１７の装置。
【請求項２１】　前記検出済みキー句は、意味タグでラベルされることを特徴とする請求項１７の装置。
【請求項２２】　前記文仮説生成器は、前記意味タグにも基づいて前記文仮説を生成することを特徴とする請求項２１の装置。
【請求項２３】　前記文仮説生成器は、前記信頼性尺度にも基づいて前記文仮説を生成することを特徴とする請求項２２の装置。
【請求項２４】　前記文仮説生成器は、前記信頼性尺度、前記意味タグ、および前記所定の意味情報に基づいて、最も確からしい文仮説を判定することを特徴とする請求項２３の装置。
【請求項２５】　前記キー句検出器は複数の部分語モデルに基づいて動作し、前記検出済みキー句は部分語の列からなることを特徴とする請求項１７の装置。
【請求項２６】　前記部分語モデルは隠れマルコフモデルからなることを特徴とする請求項２５の装置。
【請求項２７】　前記キー句検証器は、部分語モデルの集合と、対応する反部分語モデルの集合に基づいて動作することを特徴とする請求項２５の装置。
【請求項２８】　前記部分語モデルおよび前記反部分語

モデルは隠れマルコフモデルからなることを特徴とする請求項２７の装置。
【請求項２９】　前記文仮説検証器は、文仮説に対して音響的検証を実行することを特徴とする請求項１７の装置。
【請求項３０】　前記文仮説検証器は、文仮説に対して意味的検証を実行することを特徴とする請求項１７の装置。
【請求項３１】　前記文仮説検証器は、最も確からしい１つの文仮説を選択することを特徴とする請求項１７の装置。
【請求項３２】　前記検証済み文仮説に基づいて意味フレームを生成する意味フレーム生成器をさらに有することを特徴とする請求項１７の装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、音声認識の分野に関し、特に、会話音声を理解する方法に関する。
【０００２】
【従来の技術】過去数年間、会話音声の認識および理解のためのシステムが開発され、いくつもの「実世界」の応用で評価されている。いくつかのアプローチが用いられている。第１のアプローチは決定性有限状態文法（ＦＳＧ）である。これは、簡単なタスクあるいはアプリケーションに限定されるが、ユーザの発話を受け取る（そしてそれにより認識し最終的には理解する）ものである。このようなシステムでは、認識器は、音声入力全体を、可能な（すなわち、固定した文法により許容される）単語列のいずれかに合うものを見つける（復号する）ことを試みる。
【０００３】
【発明が解決しようとする課題】実際には、このような固定した文法を使用することは、ほとんど文法的に正しい文（文法内の文）がシステムに与えられる場合には有効である。しかし、多数のユーザに適用されるほとんどの典型的な「実世界」の環境では、さまざまな発話に遭遇し、その多くはこのようなタスクベースの文法によって十分に対応することができない。このような文法的に正しくない文（文法外の文）には、例えば、無関係単語、口ごもり、反復および予想外の表現などがある。日付あるいは時刻の音声認識のような明らかに単純なサブタスクの場合でさえ、自然なユーザ発話の２０％以上が文法外となる可能性が高いことが分かっている。このような条件下で、これらの文法を用いるシステムの性能は低い。そしてこの低い性能は、試行期間中に文法を念入りに調整したにもかかわらず起こる。タスクがさらに複雑な問合せに関わる場合には状況はさらに悪くなる。このような複雑な問合せに対処する固定したタスクベースの文法を書いてから十分に調整することは、法外な量の（人間の）時間および労力を必要とすることが明らかになっている。
【０００４】上記の問題点は、固定した文レベルの文法を仮定して、その文法が入力全体に適合（マッチング）しなければならないという一様な要求条件を適用する復号の枠組みから生じている。（発話の文法外部分に適合する）「フィルタ」モデルの使用は、固定文法のほとんど従う音声サンプルには限定された成功を収めているが、固定文法に固有の基本的な問題点を解決していないために、多くの一般的な発話が認識されないままである。
【０００５】会話音声の認識および理解に対するもう１つのアプローチは、統計言語モデルの使用に関するものである。このようなモデルは、固定した所定の文法に基づくのではなく、大量のサンプルデータを用いた学習（トレーニング）の結果として統計的に生成された文法に基づく。例えば、米国政府（ＡＲＰＡ）がスポンサーとなっているＡＴＩＳ（航空旅行情報システム(Air Travel Information System)）プロジェクトは、統計言語モデルのアプローチを用いた会話音声処理に対する包括的プロジェクトである。（例えば、D. A. Dahl, "Expanding the Scope of the ATIS Task: The ATIS-3 Corpus", Proc. ARPA Human Language Technology Workshop, pp.43-48, 1994、参照。）この場合、かなりの量のデータが収集され、文法外の発話を認識する能力に関しても、統計言語モデルの使用は比較的成功を収めた。
【０００６】しかし、一般的な「実世界」のアプリケーションでは、データ収集作業自体が大量の（人間の）時間および労力を必要とするため、統計モデルをトレーニングするためにこのような大量のデータを提供することは実際的でないことが多い。ほとんどのアプリケーションでは、与えられたタスクに対して単純な２連語(bigram)言語モデルを構築するのに十分な量のデータを収集するのでさえ実際的ではない。（さらに、注意すべき点であるが、ＡＴＩＳシステムの場合、実行されたシナリオおよびデータ収集はやや人工的であり、従って、実世界の自然発話に固有の問題点を必ずしも反映していない可能性がある。）これらの理由から、「実世界」のアプリケーションに配備され試験されているほとんどの会話システムは、文法外発話を処理する能力が制限されているにもかかわらず、実際には上記のような決定性有限状態文法を使用している。
【０００７】会話音声認識問題のために考えられているもう１つのクラスのアプローチは「単語スポッティング」方式に関するものである。これは、例えば、R. C. Rose,"Keyword Detection in Conversational Speech Utterances Using Hidden Markov Model Based Continuous Speech Recognition", Computer Speech and Language, 9(9):309-333, 1995、および、H. Tsuboi and Y. Takebayashi, "A Real-time Task-oriented Speech Understanding System Using Keyword-spotting", Proc. IE

EE-ICASSP, volume 1, pp.197-200, 1992、に記載されている。これらのアプローチは、入力発話の非キーワード部分のモデル化に使用する方法に依存して２つのカテゴリーに分類される。

【０００８】その第１のカテゴリーに分類することができる単語スポッティング方式は、大語彙を認識する能力に基づくものである。この方式の例としては、J. R. Rohlicek et al., "Phonetic Training and Language Modeling for Word Spotting",Proc. IEEE-ICASSP, volume 2, pp.459-462, 1993、および、M. Weintraub, "Keyword-Spotting Using SRI's DECIPHER Large-Vocabulary Speech-RecognitionSystem", Proc. IEEE-ICASSP, volume 2, pp.463-466, 1993、に記載されているものがある。この技術は、できるだけ多くの語彙知識を組み込み、キーワードモデルとともに、多くの非キーワード語彙単語モデルをシステムで利用可能とする。しかし、この技術でも、自然音声でしばしば見られる口ごもりや自己反復のような非適格な現象を十分にモデル化することができない。すなわち、すべての一様復号アプローチに固有の問題点を解決していない。さらに、大語彙自然音声認識技術は、タスク領域が限定される場合には特に、性能および効率性に問題がある。

【０００９】単語スポッティング方式の第２のカテゴリーは、入力発話の非キーワード部分をモデル化するために用いられる部分語(subword)モデルの並列ネットワークまたは単純ガーベジモデルとともに単純（すなわち限定語彙）単語スポッティングを使用するものである。このような方式の例としては、J. G. Wilpon et al., "Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models", IEEE Trans. Acoust., Speech & Signal Process., 38(11):1870-1878, 1990、および、R. C. Rose and D. B. Paul, "A Hidden Markov ModelBased Keyword Recognition System", Proc. IEEE-ICASSP, pp.129-132, 1990、に記載されているものがある。残念ながら、ガーベジモデルも、部分語モデルの並列ネットワークも、非キーワードに合うものを見つける性能が十分ではなく、そのため、キーワードモデルが発話の無関係（すなわち非キーワード）部分と誤って合わせられることも多い。この結果、多くの「フォールスアラーム」（すなわち、キーワードの誤った「認識」）が起こる。さらに、このカテゴリーに属する既存のほとんどのシステムは、語彙に依存してキーワードモデルおよびガーベジモデルを「調整」し、それにより、部分語に基づく音声認識アプローチの利点の多くを犠牲にしている。これらの理由から、このカテゴリーの単語スポッティング方式で応用が成功しているのは、例えば数字音声認識のタスクのような非常に小さい語彙を含むタスクのみである。

【００１０】

【課題を解決するための手段】本発明の発明者が認識したところでは、ほとんどの会話音声発話（すなわち「文」）はタスクに関するあるキーワードおよび「キー句（キーフレーズ）」を含み、その認識により発話の部分的あるいは全体的な理解が可能となる一方で、発話の他の部分は実際にはタスクに関係がないので無視すべきである。（注意すべき点であるが、「文」という用語は、本明細書では、任意の単語列を意味し、そのような単語列が文法的に正しい文構造を有するかどうかは問わない。また、「キー句」という用語の使用は、本明細書では、１個以上の単語からなる列を含むものとする。すなわち、キーワードは単一の単語からなる「キー句」である。）すなわち、自由度の高い音声理解システムは、文の意味的に重要な部分を検出し無関係な部分を拒絶するアプローチに基づいて構築することができる。従来の文法的制約を緩和し、かつ、認識されるキー句の集合に特に注目することによって、例えば固定した形式的な文法に基づいて受容されるよりも多くの発話が受容される。

【００１１】そこで、本発明の実施例によれば、自由度の高い（すなわち、制約のない）音声の理解を実現するために使用可能なキー句の検出および検証の技術が実現される。具体的には、単語列（すなわち文）からなる音声発話に「多重パス」手続きが適用される。まず、例えば会話の状態に固有の句部分文法の集合に基づいて複数のキー句を検出（すなわち、認識）する。次に、これらのキー句に信頼性尺度を割り当て、その信頼性尺度をしきい値と比較することによってそれらのキー句を検証し、その結果として、検証済みキー句候補の集合を得る。次に、検証済みキー句候補を、所定の（例えばタスク固有の）意味情報に基づいて結合して文仮説を得る。文仮説は、個々のキー句信頼性尺度に基づいて生成することも可能である。最後に、これらの文仮説を検証して、検証済み文仮説を生成し、その結果、音声発話の理解を得る。

【００１２】さらに、会話ベースのシステムでは特に、大規模なタスク内の会話の特定の状態に応じて（すなわち、「サブタスク」に対して）、個別の句部分文法を使用することが可能である。例えば、会話ベースの自動車予約タスク内では、システムは、与えられた時点において、要求された車が必要となる日時を決定する必要がある。この場合、予期される応答は、時間的な情報のみを与えるものであると限定することができる。自由度の高い会話マネージャと組み合わされることにより、本発明の実施例によるシステムは、文音声を少なくとも部分的に理解することができる。さらに、会話セッションが進むうちに、必要な明確化（曖昧さの除去）を実行することも可能である。

【００１３】

【発明の実施の形態】

［はじめに］本発明の実施例によれば、会話音声の認識

および理解のためのシステムは、（例えば、非キーワード大語彙知識を用いることなく）部分語ベースの音声認識の一般的な枠組みで、無関係部分を誤って「認識」せずに、発話の重要部分を認識することによって実現される。（部分語ベースの音声認識は、当業者には周知であるが、音節、半音節あるいは音素のような単語セグメントのモデリングおよびマッチングを含む。次に、それらの単語セグメント（すなわち、部分語）の列に、語彙内の各単語をマッピングするために、辞書(lexicon)が提供される。こうして、単語に対応するモデルは、実質的に、辞書によって指定される、その単語を構成する部分語のモデルの連接からなる。）図１に、本発明の実施例による音声認識および音声発話の理解を実行する１つの例示的なシステムの図を示す。

【００１４】注意すべき点であるが、従来技術の最も重大な問題点のうちの１つは、従来の音声認識器は一般に、その結果にどのくらいの信頼性をおくことができるかが分からないことである。この理由で、図１に示した本発明の実施例によれば、認識した結果に対する仮説検定を実行し、それに信頼性尺度を割り当てる検証方法を用いる。（例えば、R. A. Sukkar et al., "A Vocabulary Independent Discriminitively Trained Method for Rejection of Non-Keywords in Subword-BasedSpeech Recognition", Proc. EuroSpeech-95, pp.1629-1632, 1995、R. A. Sukkar et al., "Utterance Verification of Keyword Strings Using Word-Based Minimum Verification Error (WB-MVE) Training", Proc. IEEE-ICASSP, pp.518-521, 1996、および、M. Rahim et al., "Discriminitive Utterance Verification Using Minimum String Verification Error (MSVE) Training", Proc. IEEE-ICASSP, 1996、参照。）このような発話検証法を図１の実施例のシステムに統合することによって、キーワード（あるいは、この場合にはキー句）の検出の信頼性を高くすることができる。すなわち、キーワードモデルへの正しくないマッチングすなわち「フォールスアラーム」は大幅に減少する。

【００１５】また、図１の実施例のシステムは、このような「フォールスアラーム」をさらに減少させる。システムは、このようなキーワード（あるいはキー句）マッチングおよび検証プロセスの単独の結果として「最終判定」をしない。むしろ、検証したキーワードあるいはキー句の組み合わせ（すなわち、文）に基づいて意味解析を実行して文仮説を生成し、それを別の検証プロセスで検証する。特に、この文仮説検証プロセスは、全発話内にあるいくつかの部分語からなる「部分入力」で実行される。

【００１６】既に指摘したように、図１の実施例のシステムは、検出単位として、キーワードのみを使用するのではなく、キー句を使用する。上記の単語スポッティング方式は一般に、局所的ノイズや乱雑な音によって容易にトリガされる小さいテンプレートを使用する。より長い検出単位（すなわち、単なるキーワードの代わりにキー句）を使用することは、より特徴的な情報を含むことになり、その結果、認識段階および検証段階の両方で、より安定な音響マッチングが得られるので、有効である。

【００１７】具体的には、キー句は、１個以上のキーワードと、おそらくは、機能語との列からなる。例えば、"in the morning"は、期間についてのキー句であり、"in downtown Chicago"は、地理的場所についてのキー句である。このような句は、自然音声で発話されるときでも、一般に息継ぎなしで発話される。

【００１８】ここに記載する本発明の実施例によれば、検出されたキー句には概念情報のタグが付けられる。実際には、キー句は、例えば時刻および場所のような、意味（セマンティック）フレームにおける意味スロットに直接対応するように定義される。（意味フレームは、当業者に周知の用語であるが、与えられたアプリケーションに対して、会話によって部分的にあるいは完全に充填される情報テンプレートからなる。）従来のｎ連語(n-gram)言語モデルによって定義されるようなボトムアップ句（例えば、B. Suhm and A. Waibel, "Towards Better Language Models for Spontaneous Speech", Proc. ICSLP, pp.831-834, 1994、E. P. Giachin, "Phrase Bigrams for Continuous Speech Recognition", Proc. IEEE-ICASSP,pp.225-228, 1995、および、S. Deligne and F. Bimbot, "Language Modelingby Variable Length Sequences: Theoretical Formulation and Evaluation of Multigrams", Proc. IEEE-ICASSP, pp.169-172, 1995、参照。）とは異なり、本実施例によって認識されるトップダウンキー句は、容易に意味表現へと直接にマッピングされる。従って、これらのキー句の検出は、直接に、発話の確実な理解につながる。

【００１９】具体的には、図１の実施例のシステムは、キー句検出器１１、キー句検証器１２、文仮説生成器１３および文仮説検証器１４を有する。特に、キー句検出器１１は、会話状態に特有の句部分文法（すなわち、キー句文法２１）の集合を用いてキー句の集合を認識するための部分語ベースの音声認識器からなる。検出されたキー句には、次に、意味（セマンティック）タグが付けられる。このタグは、文仮説生成器１３（後述）によってその後に実行される文レベルの解析で有用となる。キー句検出器１１によって用いられる部分語モデル認識器は、辞書２３および部分語モデル２２を使用する。これらは、例えば、当業者に周知の従来の最小分類誤差（ＭＣＥ(minimum classification error)）基準に基づいてトレーニングされたものである。これらのモデル自体は、例えば、同じく当業者に周知の隠れマルコフモデル（ＨＭＭ）からなることも可能である。

【００２０】次に、検出されたキー句は、キー句検証器

１２によって検証され、信頼性尺度が割り当てられる。上記のように、このプロセスは、これがなければ起こり得る多くのフォールスアラームを除去する。実施例のキー句検証器１２は、当業者に周知の「反部分語モデル」を用いて、認識されたキー句の各部分語をテストする、部分語レベルの検証の組合せからなる。キー句検証器１２は辞書２３、部分語モデル２２および反部分語モデル２４を使用する。これらは、例えば、最小検証誤差（ＭＶＥ(minimum verification error)）基準を用いてトレーニングされたものである。

【００２１】図１の実施例の第３の構成要素は文仮説生成器１３である。これは、例えばタスク固有の意味情報２５を用いて、検証されたキー句候補を１つ以上の文仮説へと結合する。例えば、T. Kawahara et al., "Concept-Based Phrase Spotting Approach for Spontaneous Speech Understanding", Proc. IEEE-ICASSP, pp.291-294, 1996、に記載されたようなスタック復号器を用いて、意味制約を満たす最適な仮説を探索することができる。

【００２２】最後に、文仮説検証器１４によって、音響的かつ意味的に最良の意味仮説が検証され、最終出力（すなわち、少なくとも１つの検証された文仮説）が生成される。文仮説検証器１４は、意味情報２５、辞書２３、部分語モデル２６および反部分語モデル２７を使用する。キー句に付けられた意味タグが、キー句検出器１１によって提供され意味仮説生成器１３によって使用されるため、検証された文仮説は本質的に、直接に対応する「意味」を有し、それにより、個々のアプリケーションによる必要に応じた意味フレームの生成が可能となる。

【００２３】［キー句検出］キー句検出器１１は、キー句検出を実行する。これは、会話状態に依存する特定のサブタスクに基づくことが可能である。具体的には、各サブタスクごとに、キー句パターンが１つ以上の決定性有限状態文法として記述される。これは、実施例では、キー句検出器１１によってキー句文法２１から選択される。これらの文法は、タスク仕様から直接に人手により導出することも可能であり、あるいは、当業者に周知の従来の学習手続きを用いて、小さいコーパスから自動的または半自動的に（すなわち、人間の支援のもとで）生成することも可能である。

【００２４】一般に、キー句は、従来のキーワードに加えて、"at the"や"near"のような機能語を含む。これにより、従来のキーワードのみのマッチングに比べて、より安定なマッチングが可能となり、検出精度が改善される。（例えば、前掲のT. Kawahara et al., "Concept-Based Phrase Spotting Approach for Spontaneous Speech Understanding"を参照。）いずれのキー句にも含まれないがしばしばキー句に伴う充填句も定義され、埋め込まれたキー句を含む句パターンを形成するために使用される。

【００２５】特に、キー句および充填句の文法はネットワークへとコンパイルされる。このネットワークにおいて、キー句は繰り返し現れ、ガーベジモデルがキー句の出現の間に埋め込まれる。しかし、注意すべき点であるが、単純な繰り返しは曖昧さを生じる可能性がある。例えば、日の繰り返しが許容される場合、"twenty four"と"twenty"+"four"を区別することはできない。従って、不可能なキー句の結合を禁止する追加の制約も組み込む必要がある。

【００２６】従って、検出ユニットは、許容される結合および反復を有するキー句部分文法オートマトンのネットワークからなる。このようなオートマトンは、結合重みを評価することによって、確率的言語モデルへと容易に拡張することができる。このようなモデルを使用することにより、文レベルの文法と比べてあまり複雑にならずに、適用範囲が広くなる。

【００２７】例として、図２に、単純化した（すなわち、簡略化した）句ネットワークの例を示す。これは、「データ取得」サブタスクに適用された場合に、図１の実施例のシステムのキー句検出器１１によって使用されることが可能である。このネットワーク例の完全な実現により、曜日、月、日、および年の実質的に任意の反復が、適当な制約のもとに許容される。（このような完全な実現の全語彙は９９語である。）この特定のサブタスクでは、キャリア句は組み込まれない。

【００２８】さらに具体的には、ここに記載する本発明の実施例によって採用されている検出方法は、フォワード－バックワード２パス探索に基づくものである。これは、例えば、W. Chou et al., "An Algorithm of High Resolution and Efficient Multiple String Hypothesization for Continuous Speech Recognition Using Inter-Word Models", Proc. IEEE-ICASSP, volume 2, pp.153-156, 1994、に記載されている。本発明の別の実施例では、代わりに、当業者に周知の１パス検出法を使用することも可能である。

【００２９】A*認容スタック復号器（例えば、前掲のT. Kawahara et al., "Concept-Based Phrase Spotting Approach for Spontaneous Speech Understanding"に記載されているもの）は、Ｎ番目までの最良ストリング仮説からなる集合を求めることができるが、この結果として得られるＮ個の最良仮説は一般に、１～２個が置き換わった類似の単語列である。本発明の目標は、（入力発話全体に基づいてストリング仮説を生成することではなく）入力発話の一部に基づいてキー句候補を識別することであるので、仮説を延長しても既に延長された仮説と同じ仮説になる場合にはその仮説は捨てられる。

【００３０】特に、本実施例のスタック復号器は、キー句ネットワークのマージング(merging)状態にマークを付けることによって実現される。当業者には周知のよう

に、マージング状態は、キー句あるいは充填句が終了し、さらに延長すると次の（すなわち新たな）句の最初に侵入することになるノードに対応する。

【００３１】スタック復号器によって「ポップ」される仮説に、出力されるべき完全な句であるというタグが付いている場合、本発明の手続きは、もう１語だけその句を延長し、その句を最良延長と並べる。このノードに、以前のいずれかの仮説が同じ時点に到達している場合、検出した句を出力した後に現在の仮説は捨てられる。そうでなければ、その時点は、その後の探索のためにマークされる。

【００３２】注意すべき点であるが、この検出手続きは、冗長な仮説延長のない効率的なものであり、スコア順に、正しいＮ番目までの最良のキー句候補を生成する。本発明のさまざまな実施例によれば、この手続きは、所望の個数の句を生成したことに基づいて、あるいは、あるスコアしきい値に基づいて、終了することも可能である。例えば、仮説のスコアが、最高スコア仮説の０．９９倍より小さい値に到達したときに、検出を終了することも可能である。

【００３３】［キー句検証と信頼性尺度］図１の実施例のシステムのキー句検証器１２は、部分語レベルのテストに基づいて、検出された句の検証を行う。具体的には、与えられた句の各部分語ｎに対して、検証スコアは、次式のような従来の尤度比（ＬＲ(likelihood ratio)）テストに基づいて計算される。

$$LR_n = P(O \mid \lambda_n^c) / P(O \mid \lambda_n^a) \quad (1)$$

ただし、Ｏは、観測フレームの列を表し、λ_n^cおよびλ_n^aは、それぞれ、部分語ｎに対する正しい部分語モデルおよび反部分語モデルを表す。（部分語モデルは部分語モデル２２から得られ、対応する反部分語モデルは反部分語モデル２４から得られる。）認識の結果として、観測列Ｏは、部分語ｎに対して、ビタビアルゴリズムにより並べられ、確率$P(O \mid \lambda_n^c)$および$P(O \mid \lambda_n^a)$が得られる。（ビタビアルゴリズムは、当業者に周知の従来のスコアリング方法である。）

【００３４】各部分語モデルに対して、対応する反部分語モデルは、混同しやすい部分語クラスをまとめること（クラスタ化）によって構成される。各反部分語モデルは、対応する部分語モデルと同じ構造、すなわち、同じ個数の状態およびミクスチャを有する。反部分語モデルは、特定の部分語の検証専用であるため、反部分語モデルをリファレンスとして使用して復号を行うことにより、部分語モデルの無制約な復号を行うのに比べて、弁別性が改善される。こうして、システムは、認識器によってなされる置換誤りを拒絶する能力が増大する。この（検証）ステップでは、文脈独立な反部分語モデルを使用することも可能であるが、認識ステップは、文脈依存の部分語モデルを用いて実行される。

【００３５】特に、上記の式（１）の対数をとり、その結果を、観測Ｏの継続時間長l_nに基づいて正規化することにより、量$\log LR_n$が次のように定義される。

$$\log LR_n = (\log P(O \mid \lambda_n^c) - \log P(O \mid \lambda_n^a)) / l_n \quad (2)$$

注意すべき点であるが、式（２）の第１項は認識スコアそのものであるので、上記の計算の効果は単に、計算されるスコアを反部分語モデルのスコアだけずらし、その結果を正規化することである。

【００３６】キー句検証器１２は、検出された各キー句ごとに、対応する部分語レベルの検証スコアを組み合わせることによって、信頼性尺度（ＣＭ(confidence measure)）を計算する。例えば、検出されたキー句がＮ個の部分語を含む場合、このキー句に対する信頼性尺度は、対応するＮ個の尤度比の関数とすることが可能である。具体的には、次のようになる。

$$CM = f(\log LR_1, \ldots, \log LR_N) \quad (3)$$

信頼性尺度（ＣＭ）が、ある所定のしきい値を超える場合に、与えられたキー句は承認される。実施例では、しきい値の値は、例えば－０．１５に設定される。

【００３７】本発明のさまざまな実施例において、さまざまな信頼性尺度関数を使用することができる。例えば、第１の例示的な信頼性尺度CM_1は、フレーム継続時間による正規化に基づく。特に、これは、正しい部分語モデルに対して得られるビタビスコアと、対応する反部分語モデルに対して得られるビタビスコアの差に等しい。すなわち、次のようになる。

【数１】

$$CM_1 = \frac{1}{L}\sum_n (l_n \cdot \log LR_n) \quad (4)$$

上記の式で、l_nは、部分語ｎの継続時間を表し、Ｌは句の全継続時間である。すなわち、$L = \Sigma l_n$である。

【００３８】第２の例示的な信頼性尺度CM_2は、部分語セグメントによる正規化に基づく。特に、これは、与えられたキー句のすべての部分語の対数尤度比の単なる平均である。（一実施例では、句セグメンテーション後に単語間文脈情報が失われるため、最後の部分語に対して特別の考慮がなされる。）すなわち、次のようになる。

【数２】

$$CM_2 = \frac{1}{N}\sum_n \log LR_n \qquad (5)$$

【００３９】第３の例示的な信頼性尺度ＣＭ$_3$は、すべての部分語にわたる平均の信頼性レベルではなく、検証プロセスの結果、信頼性レベルが低いような部分語に注目する。これが有効なのは、正しくないキー句の部分語のうちには実際に入力句に正しく一致するが、他の部分語は入力句とは非常に異なることがあるからである。例えば、"November"の後半部分は、場合によって、入力句"December"の後半と完全に一致するため、部分語スコアを平均した場合に高い検証スコア（すなわち信頼性尺度）を受け取ることになる。従って、これを確実に拒絶するためには、この句の前半（その検証スコアは低くなる可能性が高い）に注目するのが有効である。

【００４０】このように、低い信頼性レベルの部分語に注目するために、各部分語ごとに正規分布を仮定することによって、対数尤度比を調整することが可能である。具体的には、部分語ＨＭＭのトレーニングで用いたサンプルを使用して、各部分語ごとに対数尤度比の平均および分散を計算する。その後、対数尤度比が、期待される平均より小さい部分語のみを含む和を実行することによって、ＣＭ$_3$を計算することができる。すなわち、次のようになる。

【数３】

$$CM_3 = \frac{1}{N_a}\sum_n \begin{cases} \log LR_n & \log LR_n < 0 \text{ の場合} \\ 0 & \text{その他の場合} \end{cases} \qquad (6)$$

ただし、Ｎ$_a$は、対数尤度比が実際には期待される平均より小さい部分語の数（すなわち、ｌｏｇ ＬＲ$_n$＜０となる部分語の数）である。

【００４１】第４の例示的な信頼性尺度ＣＭ$_4$はシグモイド関数を用いる。この例示的な信頼性尺度は、最小誤り率基準でトレーニングするための損失関数として用いられる。すなわち、次のようになる。

【数４】

$$CM_4 = \frac{1}{N}\sum_n \frac{1}{1+\exp(-\alpha \cdot \log LR_n)} \qquad (7)$$

これらの信頼性尺度のそれぞれに対して（あるいは、本発明の別の実施例によって使用される信頼性尺度に対して）、特定のしきい値を選択することが可能である。与えられた信頼性尺度の値がそのしきい値を下回る場合、候補キー句は検証済みキー句候補の集合から排除され、そうでない場合、検証済みキー句候補の集合に含まれる。

【００４２】本発明のさまざまな実施例によれば、計算される信頼性尺度の尤度比は、「フォールスアラーム」を排除するためだけではなく、検証済みの句に対する「再スコアリング」を行うための基礎としても使用可能である。例えば、E. Lleida and R. C. Rose, "Efficient Decoding and Training Procedures for UtteranceVerification in Continuous Speech Recognition", Proc. IEEE-ICASSP, pp.507-510, 1996、には、尤度比に基づいて復号を行うことが提案されている。しかし、尤度比の直接の使用は、そのダイナミックレンジが大きいため、不安定となる可能性がある。こうして、本発明の一実施例によれば、反部分語モデルのスコアが正しい部分語モデルのスコアより大きい場合（すなわち、ＣＭ$_1$＜０の場合）にガーベジ充填句を生成することによって、反部分語モデルをガーベジモデルとして処理する。ガーベジ充填句は、もとの句と同じ継続時間を有し、もとの句よりも例えばＣＭ$_1$だけ高いスコアを有する。その結果、もとの句は、その後の文解析（以下参照）で選択される可能性が低くなる。

【００４３】［文解析］図１の実施例のシステムの文仮説生成器１３はキー句検証器１２によって生成された検証済みキー句候補を意味情報２５に基づいて１個以上の文仮説へと組み合わせる文解析を実行する。一実施例では、句候補のＬＲ(left-to-right)トレリスを使用することが可能な１次元ＲＬ(right-to-left)探索が用いられる。別の実施例では、島駆動探索アルゴリズムを用いることも可能である。トレリス解析は計算量が多いため、さらに別の実施例ではラティス解析法を採用する。これは、トレリス解析よりわずかに精度が低くなるだけである。ラティス解析法は、音響スコアと、提供される意味制約情報（キー句タグの許容される組み合わせを指定する）に基づいて、句候補を結合する。キー句検出のためのフォワードーバックワード探索によって与えられるスコアを音響スコアとして用いることが可能である。

【００４４】最も可能性の高い文仮説を効率的に見つけるためには、スタック復号探索法を採用すると有効である。この方法は、一連の部分仮説を反復的に生成し、完全な文仮説が生成されるまで、各反復において最良の利用可能な部分仮説を延長する。

【００４５】具体的には、現在の「最良の」部分仮説を｛ｗ$_1$，ｗ$_2$｝とし、新たな仮説が句ｗ$_3$を連結することによって生成されると仮定する。新たな仮説｛ｗ$_1$，ｗ$_2$，ｗ$_3$｝に対する評価関数は、完全な入力発話ｈ$_0$に対する上限スコアからのずれ（オフセット）として以下のように計算される。

【数５】

$$\hat{f}(w_1, w_2, w_3) = h_0 - (h_0 - \hat{f}(w_1)) - (h_0 - \hat{f}(w_2)) - (h_0 - \hat{f}(w_3))$$
$$= \hat{f}(w_1, w_2) - (h_0 - \hat{f}(w_3))$$

ただし、ｆ＾（w_i）は、検出された句w_iに対する評価の結果である。初期仮説はｆ＾（null）＝h_0である。新たな句が追加されるごとに、オフセットが減算される。上限h_0は、認識プロセスのフォワードパスで計算される。

【００４６】上記の方法は、例えば、W. A. Woods, "Optimal Search Strategies for Speech Understanding Control", Artificial Intelligence, 18:295-326, 1982、に記載されているような不足法(short-fall method)に基づいている。注意すべき点であるが、この評価はA_*認容である。しかし、探索を効率的に導くこの方法の発見的能力はやや限定されたものとなる可能性がある。検出ベースの解析段階では特に、入力発話全体が扱われることを仮定しないため、数語の短い仮説が誤って受容される可能性が高い。従って、発話でスキップされた部分を評価することが有効となる。そのため、具体的には、本発明の一実施例によれば、スキップ長に比例する一様な罰金値をオフセットとして追加することが可能である。もちろん、この近似は粗雑であるため、次善の探索となる可能性がある。従って、これを補うために、できるだけ多くのキー句とともに、できるだけ多くのガーベジ句（無音を含む）を生成することが好ましい。（一実施例では、これらの仮説は、キー句検証プロセス中に生成することも可能である。）

【００４７】［文検証］図１の実施例のシステムの文仮説検証器１４は、認識出力の最終判定を行う。実施例では、大域的音響情報および大域的意味情報の両方を使用し、それぞれ入力発話全体に適用される。キー句検証プロセスは局所的な判定のみをしたが、文仮説検証プロセスはこれらの局所的な結果を組み合わせ、従来の発話検証と同様の効果を実現する。しかし、検出ベースの認識プロセスは、多数の予期しないキャリア句を含む場合でも入力発話を受容することが多いことに注意すべきである。

【００４８】具体的には、文仮説検証器１４によって実行される音響検証プロセスは、与えられた文仮説が十分に一致することを保証するために、入力発話全体の再スコアリングを行う。この再スコアリングは、部分語モデル２６、反部分語モデル２７、および辞書２３を用いて行われる。この段階で適用される部分語モデル（すなわち、部分語モデル２６）の集合および対応する反部分語モデル（すなわち、反部分語モデル２７）の集合は、キー句検出器１１およびキー句検証器１２によって使用されるもの（すなわち、部分語モデル２２および反部分語モデル２４）よりも精度が高い。こうして、より高い精度の音響再スコアリングが実行される。

【００４９】一方、意味検証プロセスは、与えられた各文仮説の意味的「完全性」を評価する。例えば、本発明の一実施例によれば、意味検証は、ある構成要素が意味的に「合法」かどうかのみを指定する単純な意味制約情報に基づいて実行される。このような場合、文仮説検証器１４の意味解析部分は、例えば、与えられた文仮説の意味表現が完全であるかどうかを判断する。しかし、注意すべき点であるが、会話ベースのアプリケーションでは、例えば、不完全な発話にしばしば遭遇する。例えば、ユーザはただ"August"（８月）と言うだけで、その月の特定の日を指定しないことがある。一般に、こうした「不完全な」発話も同様に受容すべきである。

【００５０】従って、本発明の一実施例によれば、文仮説検証器１４は、与えられた文仮説が意味表現を完成しておらず、かつ、ほとんどの入力セグメントが尤度比テストで拒絶された場合にのみ、その文仮説を拒絶する。この組合せ「テスト」は、例えば、満足な文仮説に遭遇するまで、各文仮説に適用することが可能である。

【００５１】しかし、本発明の別の実施例では、さらに一般的な確率的意味モデルを、文仮説検証器１４で用いることが可能である。このような場合、各文仮説について、音響スコアとともに意味スコアを求め、組み合わせたスコアを用いて、最終認識結果として出力すべき検証された文仮説を選択することが可能である。

【００５２】本発明のさらに別の実施例では、意味的検証のみまたは音響的検証のみ（両方ではない）を、文仮説検証器１４で実行することが可能である。例えば、さらに高い精度の部分語および反部分語のモデルが利用可能でない場合には、入力発話の音響再スコアリングを実行することはあまり効果がない。従って、この場合、意味検証のみを実行して、単に、与えられた文仮説が意味表現を完成していることを検証するか、あるいは、確率的意味モデルを用いている場合には、検証済み文仮説が最終認証結果として判断されるもとになる意味スコアを生成する。

【００５３】［付記］説明を明確にするため、ここに記載した本発明の実施例は、個別の機能ブロックからなるものとして表した。これらのブロックによって表される機能は、共用あるいは専用のハードウェアの使用によって提供することが可能である。ハードウェアには、ソフトウェアを実行することが可能なハードウェアが含まれるが、これに限定されるものではない。例えば、ここに記載した構成要素の機能は、単一の共用プロセッサによって、あるいは、複数のプロセッサによって提供することが可能である。本発明の実施例は、ディジタル信号プロセッサ（ＤＳＰ）ハードウェア、上記の動作を実行するソフトウェアを格納する読み出し専用メモリ（ＲＯＭ）、および、結果を格納するランダムアクセスメモリ

（ＲＡＭ）からなることが可能である。超大規模集積（ＶＬＳＩ）ハードウェアや、カスタムＶＬＳＩ回路を汎用プロセッサやＤＳＰ回路と組み合わせたものも可能である。

【００５４】また、「キー句検出器」、「キー句検証器」、「文仮説生成器」、および「文仮説検証器」という用語は、対応する機能を実行する任意のメカニズムを含む。

【００５５】

【発明の効果】以上述べたごとく、本発明によれば、固定した形式的な文法に基づいて受容されるよりも多くの発話が受容される、効率および自由度の高い会話音声認識が実現される。

【図面の簡単な説明】

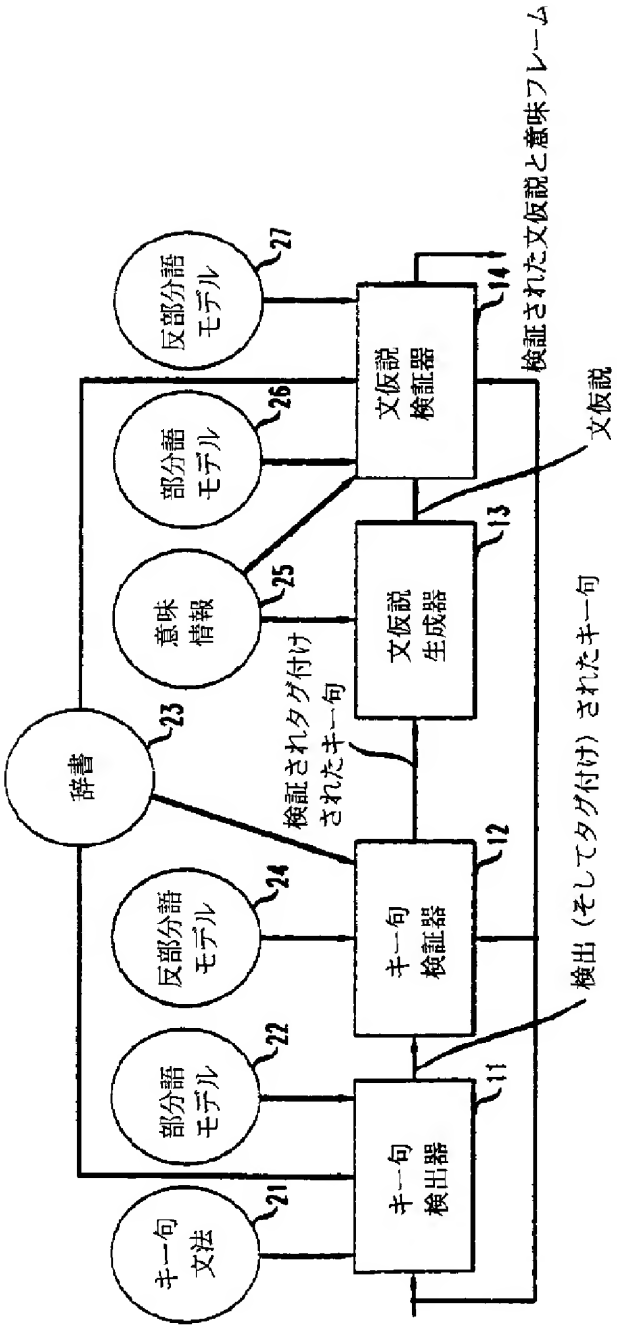
【図１】本発明の実施例による音声認識および音声発話の理解を実行するシステムの図である。

【図２】「日付取得」サブタスクに適用した場合に、図１の例示的なシステムによって使用されることが可能な単純化された句ネットワーク例の図である。

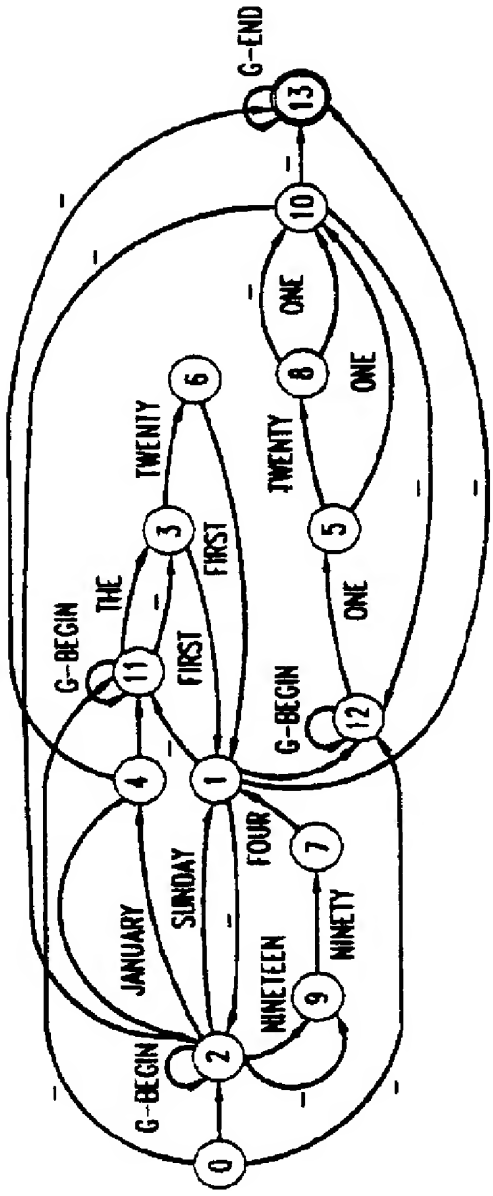
【符号の説明】

１１　キー句検出器
１２　キー句検証器
１３　文仮説生成器
１４　文仮説検証器
２１　キー句文法
２２　部分語モデル
２３　辞書
２４　反部分語モデル
２５　意味情報
２６　部分語モデル
２７　反部分語モデル

【図1】



【図2】



フロントページの続き

(71)出願人　596077259
600 Mountain Avenue, Murray Hill, New Jersey 07974－0636U. S. A.

(72)発明者　ビン－ホワン　ジャン
アメリカ合衆国、07059　ニュージャージー、ウォレン、サウス　レーン　８

(72)発明者　かわはら　たつや
京都府京都市伏見区東奉行伏見御堂122

(72)発明者　チンーフイ　リー
アメリカ合衆国、07974　ニュージャージー、ニュー　プロビデンス、ラニーメデパークウェイ　118